November 12, 2010

Linda S. Birnbaum, Ph.D., D.A.B.T., A.T.S.
Director
National Institute of Environmental Health Sciences
& National Toxicology Program
P.O. Box 12233
Research Triangle Park, North Carolina 27709

Re: NTP Evaluation and Classification of Formaldehyde for the 12th Report on Carcinogens

Dear Director Birnbaum:

in our September 7th letter to Dr. Collins, and in our September 27th letter to the National Academy of Sciences (NAS) panel reviewing formaldehyde(Attachment A), we expressed our concerns with the National Toxicology Program's (NTP) conclusion that formaldehyde should be classified in the 12th Report on Carcinogens (RoC) as a chemical "known to cause" myeloid leukemia. The conclusion reached by the NTP is largely based on reports of workers in industry (Hauptmann et al. 2003; Pinkerton et al. 2004; Beane Freeman et al. 2009) and embalmers (Hauptmann et al. 2009).

We are writing to you to provide new, relevant information and data/analyses (Attachment B) of the NCI's study of embalmers reported by Hauptmann et al. (2009) in response to your letter to us dated October 5, 2010 that included an "informational document" in which it was stated that NTP would consider "new, relevant scientific information" . The attached new analyses were not considered by the NTP prior to the end of the comment period in June, 2010, nor were they adequately addressed by the study authors (Hauptmann et al. 2010) in response to our letter to the Journal of the National Cancer Institute (JNCI) in which we raised these concerns (Cole et al. 2010) (Attachment C). We have summarized these analyses in the following paragraphs.

Incomplete and Misleading Interpretation of the Statistical Analyses

Our letter to JNCI (Cole et al. 2010) noted that the Odds Rations (ORs) on exposure-response in Table 4 of the Hauptmann et al. (2009) study were not tested for statistical significance. **This is critically important for the following reasons**:

- The Hauptmann et al. (2009) paper present p-values derived from the continuous (unconditional logistic regression) analyses for the entire population, and those p-values are reported in both Tables 3 and 4.
- These p-values are misleadingly juxtaposed with the ORs derived from the categorical analyses presented in Tables 3 and 4, which comprise the only descriptive results in this

paper, and have not themselves been tested for statistical significance in either table. To illustrate:

- Consider the results in Table 4 for "peak" exposure and myeloid leukemia. The ORs for increasing levels of peak exposure to formaldehyde are 2.9, 2.0 and 2.9, and the seemingly related p-value (excluding the crucial reference group with an OR=1.0) is an *inverse,* non-significant -0.778.
- When the reference group is included, the "continuous variable" test gives a positive result (p=0.036); however, this test is not based on the categorical ORs.
- In any case, and no matter how it is assessed, there is no exposure-response relationship among the critical groups, the exposed subjects themselves.

The combination just mentioned – i.e., the statistically non-significant inverse trend and the statistically significant positive trend seen in the continuous data and an essentially flat pattern of categorical ORs among exposed subjects - discussed in greater detail in our Attachment B, could have arisen because of an underlying *non-linear* relationship between logit p and continuous exposure, which was apparently not considered by the authors.

In addition to the continuous model estimates, the authors should have presented trend tests based on the midpoints (or scores) of the categorical data that they presented. The arbitrariness of the categories notwithstanding, trend tests based on their midpoints (which also require underlying linearity, here with respect to logit p vs. midpoint score) would directly reflect the apparent pattern of categorical ORs presented in the tables and avoid misinterpretation or confusion about trends.

Significant Uncertainties in the Estimates of Exposure

We have raised our concerns (Attachment A) with the estimates of exposure used as the basis for the conclusions drawn by Hauptmann et al. (2009), which were not considered by the NTP prior to June 2010, to include:

- Information on important variables (e.g., time to perform an embalming, years of embalming) was typically missing for 35% to 45% of subjects.
- When more than one interview was available for a subject, information was discordant in 34% to 43% of comparisons, depending on the index. The manner of resolving these discrepancies was not described.
- About 44% of embalmers began embalming before 1933, and 76% began before 1943. All embalmers included in this study died before 1986. But, interviews were not conducted until 1990-92 – i.e., up to 32 years after the death of a subject, and up to 60 (and even more in a few instances) years after subjects began embalming.

- Peak exposure to formaldehyde (one of the two exposure indices reported to be significantly associated with mortality from myeloid leukemia) "could not be validated..." by the exposure experiment.

Lack of Consideration of the Results of Sensitivity Analyses on Conclusions Reported by Hauptmann et al. (2009)

These analyses were done by Hauptmann et al. (2009) to assess the effects of missing data. The results were reported for five exposure indices after the exclusion of subjects whose work histories were less than 70% complete. The results of the refined sensitivity analyses were profoundly different from the published results based on all subjects, which required a large amount of "imputed" data. For example:

- The average OR, calculated by us, for the highest exposure level of the five indices was reduced from 3.6 in the published full data set to 1.9 in the refined data.
    - The effect on peak exposure is especially notable: the OR was reduced from 3.8 (1.1-12.7) to 1.2 (0.3-5.3), that is, was no longer statistically significant.
    - When we focused not on the ORs but on the "effect measure" (OR-1), the value declined from 2.8 to 0.2 and became essentially null.
    - Even for "number of embalmings", one of the variables in which the authors of this study"...have the most confidence", the OR was reduced from 3.9 to 2.3 and the confidence interval included the null, that is, was no longer statistically not significant.

Interpretation of Our Additional Analyses of the Hauptmann et al. (2009) Study

Based on the above summary and the details given in Attachment B, our view is that:

- The information from the interviews is incomplete, and that which is available is of questionable validity.
- The decline in the ORs seen in the sensitivity analyses are attributed by Hauptmann et al. (2009) to, "...smaller numbers of subjects and to chance", although these are largely one and the same. Perhaps chance was a factor in lowering the Ors, but the sensitivity analyses indicate that the process of imputation (replacing a subject's missing value for an exposure index with an estimate derived from similar subjects) exerted a different effect on the data of the controls than on the data of the cases. This effect was both substantial and systematic, leading to considerable bias, i.e., over-estimation, of the ORs .
- Finally, the reported findings for the indices that are based in part on the exposure experiment, particularly the estimates of peak exposure, are so unreliable as to be uninterpretable.

Further, the results of the Beane Freeman study did not demonstrate a significant increase in myeloid leukemia in workers by any dose metric considered. Also, when under-reported deaths were considered, a reanalysis of the Hauptmann et al. (2003) study of the same industrial population did not show a statistically significant increase between formaldehyde exposure and myeloid leukemia (Beane Freeman et al. 2009 – supplemental analyses; Marsh et al. 2004). Further, the Pinkerton et al. (2004) study that showed an increase in all myeloid leukemia, but not acute myeloid leukemia, did so only with one index of exposure. The results of these epidemiological studies upon which the NTP has based their conclusion do not rise to the level of evidence required to demonstrate a causal association between formaldehyde exposure and myeloid leukemia in those workers.

We also described in our October 29th letter to you new relevant data for each of the workers included in the Zhang et al. (2010) study; data that were not included in the published paper and also not considered by the NTP. We received these data in July 2010 in response to a Freedom of Information Act request. These new data and our analyses of them clearly demonstrate that the conclusions drawn by Zhang et al. may not be used to support a biologically plausible link between formaldehyde exposure and myeloid leukemia; rather, the study suffered from methodological deficiencies, misinterpretations of results, and a serious lack of scientific foundation for its conclusions. Our letters to you (October 29, 2010) and to the NAS (October 4, 2010) outlining these issues are attached (Attachments D and E, respectively).

Further, issues with regard to the mode of action for formaldehyde-induced myeloid leukemia also have been published since June 2010, specifically with regard to the Zhang et al. study (Goldstein 2010), and more generally concerning and with formaldehyde's mechanism of action in the production of leukemia (Ward et al. 2010). Of particular note, the latter paper prepared by scientists from academia and regulatory and advisory agencies including IARC, NCI, USEPA, and ATSDR (Ward et al. 2010) states that,

> …more research is needed to elucidate the mechanism by which formaldehyde could cause myeloid leukemia in humans.

Goldstein (2010) stated that:

Replication of the findings of Zhang et al. remains central to the question of whether formaldehyde should be considered to be a known rather than a probable human leukemogen – particularly without a clear understanding of the mechanism by which inhaled formaldehyde reaches bone marrow stem cells without further independent replication of the epidemiological association of formaldehyde exposure with leukemia.…[T]he present evidence is not quite

sufficiently strong to warrant the designation given by IARC and by other review organizations of [formaldehyde] being a known human leukemogen.

We understand that there is no legal deadline for NTP to complete its review and decision-making regarding formaldehyde. Consequently, NTP has the opportunity to fully evaluate and consider the scientific value of the information and analyses provided in our September 7and October 29 letters and those contained in the Attachments to this letter prior to the completion of the 12th RoC documents that are made available to the BSC, the NTP Executive Committee, and HHS Secretary Sebelius. Ignoring these critical parts of the body of scientific evidence will not serve science or public health.

Based upon your October 5th letter, to us, the NTP had not yet completed its review process for the candidate substances, including formaldehyde, that are under consideration for listing in the 12th RoC. Moreover your October 5th letter states that NTP will consider "new, relevant scientific information on formaldehyde". We have presented such new, relevant information and request that this new information and analyses be fully considered and directly addressed by the NTP.

In light of this critical new information and these analyses – both in addition to and in advance of your informing the Secretary Sebelius "of the points that [we] raise" in our September 7th letter to Dr. Collins, our letter to you on October 29th and the enclosed Attachments, we request that NTP carefully and thoroughly evaluate and address the important scientific matters contained in this letter and Attachments.

Please contact Dr. Kenneth Mundt (at 413-256-3556) with any questions or comments, or otherwise wish to discuss these matters. Thank you.

Kenneth A. Mundt, PhD
Principal and Director of Epidemiology
ENVIRON International Corporation
Amherst, MA

Philip Cole, MD
Professor Emeritus of Epidemiology
University of Alabama at Birmingham
Birmingham, AL

Richard Irons, PhD
Professor Emeritus
University of Colorado at Denver
Cinpathogen International, LLC
Boulder, CO

Gary M. Marsh, Ph.D., F.A.C.E.
Professor of Biostatistics, Epidemiology and Clinical & Translational Science
Director, Center for Occupational Biostatistics & Epidemiology
Department of Biostatistics
Graduate School of Public Health
University of Pittsburg
Pittsburgh, PA

Jack S. Mandel PhD, MPH
Chief Science Officer
Exponent
Menlo Park, CA

Enclosures

References:

Beane Freeman L, Blair A, Lubin J, et al. (2009). Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: update of the NCI cohort. ***J Natl Cancer Inst.*** 101(10):751-761.

Cole P, Adami H-O, Trichopoulos D, Mandel J. (2010). Lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. ***J Natl Cancer Inst***, online.

Hauptmann, M, Lubin, J, Steward, P, et al. (2003). Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries***. J Natl Cancer Inst***. 95(21): 1615-

Hauptmann M, Stewart P, Lubin J, et al. (2009). Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. ***J Natl Cancer Inst***. 101(24):1696-1708.

Hauptmann M, Stewart P, Lubin J, et al. (2010). Response. ***J Natl Cancer Inst***, online.

Marsh, G and Youk A. (2004). Reevaluation of mortality risks from leukemia in the formaldehyde cohort study of the National Cancer Institute. ***Reg Tox and Pharm.*** 40(2):113-24.

Pinkerton, I, Hein, M, Stayner, L. (2004). Mortality among a cohort of garment workers exposed to formaldehyde: an update. ***Occup Environ Med.*** 61:193-200.

Zhang, L, Tang, X; Rothman, N, et al. (2010). Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells***. Cancer Epidemiology, Biomarkers & Prevention*** 19(1):80-88.

Attachment A:
Letter to the National Academy of Sciences
Epidemiological Issues

September 27, 2010
Dr. Ellen Mantus
Senior Program Officer for Risk Analysis
National Academy of Sciences
500 Fifth St. N.W.
Washington, D.C. 20001
Email: emantus@nas.edu
RE: **Summary of Key Concerns Regarding the Epidemiological Studies Reported in the IRIS Draft Formaldehyde Document**

Dear Members of the NAS Committee:
The undersigned prepared comments which were presented to your Committee by some of on August 9, 2010, during the allocated public comment period. Collectively, we expressed several concerns regarding the IRIS draft document, specifically pertaining to formaldehyde as a leukemogen. We understand and appreciate that the Committee has been evaluating the draft IRIS document, and has considered the points we raised. However, because the five minutes we each were given was too short to articulate all key epidemiological points, we are providing this brief summary for your consideration.

IRIS review methodological issues

1. The standard criteria used to critically review individual epidemiological studies are not transparent.
2. Inconsistencies in results identified across studies appear not to be addressed in the synthesis of evidence, and therefore not reflected in the conclusions.
3. The methodology for synthesizing epidemiological evidence across relevant studies is not clear.
4. Although strengths and weaknesses of the various articles are discussed, the criteria for determining relative study weights are not transparent.
5. Overlapping cohorts and studies are not properly identified and discounted.
6. Some papers have been overlooked or excluded without justification.
7. There is a lack of consistency across exposure metrics (including contradictory evidence such as 'ever peak' vs. 'cumulative peak' exposures) within and across studies.
8. The consequences of combining evidence across studies with non-comparable exposures and inconsistent exposure metrics have not been addressed.
9. Summary tables depicting evidence on which conclusions are based are lacking.

10. Causal conclusions are offered for combinations of unrelated diseases of differing etiologies – e.g., all lymphohematopoietic cancers combined, all leukemias combined and all myeloid leukemias combined.
11. Undue weight is placed on selected studies of embalmers (i.e., three PMR studies and case-control study based on these cases), plus the Zhang meta-analysis.
12. Preferential identification of "positive" associations or "increased" risks (many not statistically significant) is relied upon in formulating causal conclusions.
13. Perhaps most crucial is that EPA's causal conclusions do not clearly follow from their own critical review of the individual studies.
14. Ultimately, the IRIS draft does not appreciate that the body of study findings is predominantly negative when a proper "weighting" of the evidence is considered. It is especially striking that the strong evidence of no increased risk based on the three largest cohort studies combined (152 observed and 153 expected leukemia cases) is afforded little weight and that the considerably weaker Hauptmann 2009 report is repeatedly highlighted.

<u>Critical points pertaining to influential studies in IRIS Draft</u>

Beane Freeman 2009

1. Supplemental material is published on-line that corrects the results from Hauptman 2003, including 995 (erroneously reported as 1006) previously omitted deaths. This should be highlighted.
2. No excesses of leukemia or myeloid leukemia deaths are reported in this study.
3. Quantitative exposure estimates based on industrial hygiene measurements generate no consistent exposure-response relationships. Whereas "ever peak" exposure produces a positive association, no association is seen (and results not shown) for "cumulative peak" exposure, a more accurate indicator of high exposure.
4. The percentage of statistically significant tests reported (3/36 or 8.3%) is close to what would be expected (5%) by chance alone.

Zhang 2009

5. This meta-analysis does not follow standard methods for selection of data from studies.
6. Data are combined across studies for the "highest exposed" category from each, regardless of the comparability of exposure metric.
7. Studies with overlapping (non-independent) populations have been included.
8. The meta-analysis does not include the Beane Freeman 2009 study.

Hauptmann 2009

9. This study relies on convenience sample of death certificates from several previous reports.
10. Cases were identified based on both contributing and underlying causes of death.
11. Many death certificates reflected coding practices and diagnostic criteria from previous decades and may not be comparable over time.

12. The study data do not demonstrate an excess of myeloid leukemia: PMR=108, 95% CI 72-156 (Cole 2010, in press).
13. Surrogate exposure information was obtained from next-of-kin or co-workers, often pertaining to decedents' work practices from several decades earlier and subject to recall bias.
14. Myeloid leukemia cases differed from controls in several ways, possibly reflecting selection forces: cases were first employed earlier (52% more likely employed prior to 1942) longer, and at a younger age; cases had earlier typical year of death (subject to diagnostic and coding conventions of the 1960's and 1970's); cases had higher number of embalming and estimated cumulative exposure (possibly reflecting differences in employment duration); and all cases were of white race.
15. On the other hand, myeloid leukemia cases were similar to controls with respect to the following key exposure indicators: average formaldehyde estimates; TWA 8-hour exposure estimates; and peak formaldehyde exposure estimates.
16. For myeloid leukemias, the reference group suffered from extremely small numbers, leading to two sets of analyses. Initial findings using the unexposed as referent – which only contains one myeloid death – yields highly unstable risk estimates; therefore the referent group is re-set to include those having conducted 500 or fewer embalming. Basis for selecting 500 as "unexposed" is not evident.
17. The second set of analyses, using the expanded reference group and generating more stable estimates, does not demonstrate dose-response relationships across exposure categories for most exposure surrogates.
18. Tests for trend, based on an analysis that is not presented, are juxtaposed with results from both the first and second analyses, which is misleading, as they appear to apply to the presented results.
19. Interpretation of the study OR's is difficult, since the study base does not represent a specified cohort's person-time, cases represent only those identified from available death certificates, and controls could not be randomly sampled from the actual cohort that generated cases.
20. Other case-control studies on this subject were based on registry cases (e.g., Blair 2001, Partanen 2003), with more comprehensive and verifiable diagnostic data. These studies found little evidence of an association between formaldehyde exposure and leukemias or myeloid leukemias.

<u>Omissions from IRIS Draft relevant to leukemia discussion</u>

Bachand 2010

1. This is the only meta-analysis on this topic that used standard methods and included Beane Freeman 2009.
2. More sensitivity analyses are presented than in any other meta-analysis, and the associations between formaldehyde and leukemia and myeloid leukemia are not robust.

Marsh 2004

3. This re-analysis of the data from Hauptmann 2003 demonstrates unusually low disease rates in the reference group used in internal exposure-response analyses.

4. The results bring into question the suggestion made in Hauptmann 2003 that the observed associations reflect a causal relationship.

Marsh 2010

5. This re-analysis demonstrates the impact of NCI's omission of 995 deaths (12% of all deaths) on the study results.
6. The corrected analyses show attenuated associations because the omitted observed deaths disproportionately occurred among unexposed employees.
7. The re-analysis also demonstrates latent periods that are implausibly longer than those observed related to other chemical exposures such as chemotherapeutic agents.

Lu 2010

15. The Lu 2010 study, though not an epidemiological study, helps explain the general lack of observed excesses of leukemia, especially given that formaldehyde appears to be incapable of reaching blood-forming tissues.

We appreciate your interest and consideration. Please do not hesitate to contact any of us should you have any questions or if any of these points is unclear.

Kenneth A. Mundt, PhD
Principal and Director of Epidemiology
ENVIRON International Corporation
Amherst, MA

Jack S. Mandel PhD, MPHChief Science Officer
Exponent
Menlo Park, CA

Philip Cole, MD
Professor Emeritus of Epidemiology
University of Alabama at Birmingham
Birmingham, AL

Gary M. Marsh, Ph.D., F.A.C.E.
Professor of Biostatistics, Epidemiology and
Clinical & Translational Science
Director, Center for Occupational Biostatistics
& Epidemiology
Department of Biostatistics
Graduate School of Public Health
University of Pittsburg
Pittsburgh, PA

Attachment B:
Analyses of the Hauptmann et al. (2009) Study

Mortality from Lymphohematopoietic Cancer (LHC) Among Embalmers:
Response to a Letter by Hauptmann, et al (2010).

P. Cole, H-O. Adami, D. Trichopoulos, J. Mandel, G. Marsh
October 15, 2010

In December, 2009 the *Journal of the National Cancer Institute* ***(JNCI)*** published an epidemiologic study by Hauptmann et al (1) of lymphohematopoietic cancer (LHC) among embalmers (the "embalmers study" (ES)). Estimates of formaldehyde exposure were developed from interviews and from the results of an "exposure-assessment experiment" (2). The data were presented as a case-control analysis of a proportional mortality ratio data set.

The major findings were that years of working as an embalmer and lifetime highest peak exposure to formaldehyde, were statistically significantly associated with mortality from the myeloid leukemias (MLs) (1).

In September, 2010 the ***JNCI*** published online our letter raising questions about the ES (3) and a response (4) from its authors. Citations to these letters are not yet available.

This rebuttal to the response by Hauptmann et al. has been prepared for our own use and for distribution to interested parties. Issues are addressed in the order in which they appear in the response.

I. Data Sources and Proportional Mortality Ratios

The response to our letter provides a brief history of the selection of the ES's subjects (decedent embalmers) from three proportional mortality ratio (PMR) studies (5-7) published in 1983, 1984 and 1990. (There is not a one-to-one relationship of the subjects in these early studies with those in the ES.) The response also provides PMRs reported in those early studies for leukemia (meaning "all" leukemia). These values are: 140, 175 and 152 and the underlying data yield a pooled PMR of 153(95% CI: 120-193).

It is unclear why the response provided PMRs for "all" leukemia since the ES and our letter were focused on the MLs. Further, the ES database would permit the calculation of many PMRs including that for the MLs. In fact, PMRs for the MLs are available from data in the earlier studies; they are: 146, 150 and 157 and the pooled PMR is 154 (108-213).

The response to our letter states that the PMR we provided for the MLs (108(70-156)) is "crude...and provide(s) no additional useful information". Our PMR is somewhat crude as it had to be based on data in the ES. Those data are presented in groups for subjects matched to the demographic characteristics of all LHC decedents - not to those of all embalmers in the

original studies. Our PMR could be refined by subject restriction but this would involve assumptions that may not be accurate.

II. Internal Versus External Measures of Association

The response to our letter endorses the use of internal measures of association (e.g., odds ratios, relative risks) particularly with respect to exposure-response analyses. It also diminishes the value of external measures (PMRs, most SMRs). We are in agreement with this emphasis on odds ratios (ORs) in the analysis of case-control data. Our related point was somewhat different and we do not think it is controversial. We suggest that the interpretation of ORs is enhanced when they can be related to an external measure of risk such as a PMR. For example, Table 4 in the ES indicates that about 69% (182/265, based on the control series) of embalmers had an exposure to formaldehyde sufficient to raise their OR of the MLs to about 2.5 (the average OR among exposed subjects in Table 4) – as compared to a value of 1.0 for the reference group. If the overall PMR for the embalmers is 108, as we suggest, then the OR of 2.5 among exposed subjects requires that the reference group (about 31% of embalmers) would have a very low absolute risk of the MLs. It would be equivalent to a PMR of about 53. Then the exposed groups would have a PMR of only about 133.

III. Participation Rates

Our letter stated that important information relating to the interviews conducted, including participation rates, was not provided in the ES. The response repeats the ES and states that the interview response rate was about 95%. However, this figure represents only the percentage of decedent subjects for whom some interview information was available. It is not the level of participation among the living persons (next-of-kin, co-workers) who were available, or requested, to provide an interview.

A full description of interview procedures and results, including participation rates, should allow assessment of the validity of the interview information before it is transformed into the indices (not “metrics”; the ES includes no measurements) of exposure to formaldehyde.

IV. Data Validity

With respect to the likely validity of the exposure indices we considered the following:

1. No interview was available for about 5% of all subjects.
2. Information on important variables (e.g., time to perform an embalming, years of embalming) was typically missing for an additional 35% to 45% of subjects.
3. When more than one interview was available for a subject, information was discordant in about 34% to 43% of comparisons, depending on the index. The manner of resolving these discrepancies is not described.

4. About 44% of embalmers began embalming before 1933 and 76% began before 1943. All embalmers in the ES died before 1986.
5. But, interviews were not conducted until 1990-92 – up to 32 years after the death of a subject and up to 60 years (more in a few instances) after subjects began embalming.
6. Peak exposure to formaldehyde (one of the two exposure indices reported to be significantly associated with mortality from the MLs) "could not be validated..." by the exposure experiment.

V. Sensitivity Analyses

These analyses were done in the ES to assess the effects of missing data. The results were reported for five exposure indices after the exclusion of subjects whose work histories were less than 70% complete. This restriction reduced the number of exposed cases of malignancies of non-lymphoid origin (70% of which are MLs) by about 40% (from 44 to about 27); this percentage differed from one exposure index to another. The effect on control subjects is not provided.

The results of the refined analyses - based on fewer subjects with more complete data - were profoundly different from the published results based on all subjects and including a large amount of "imputed" data. ORs were reduced considerably in all five of the sensitivity analyses that were done. The average OR, calculated by us, for the highest exposure level of the five indices was reduced from 3.6 in the published full data set to 1.9 in the refined data. The effect on peak exposure is especially notable: the OR was reduced from 3.8 (1.1-12.7) to 1.2 (0.3-5.3). If we focus not on the ORs but on the "effect measure" (OR-1), the value declines from 2.8 to 0.2 and becomes essentially null. Even for "number of embalmings", one of the variables in which the authors of the ES "...have the most confidence", the OR was reduced from 3.9 to 2.3.

Interpretations - Based on sections III and IV above our view is that the information from the interviews is very incomplete and that which is available is of questionable validity.

The decline in the ORs seen in the sensitivity analyses are attributed by the authors of the ES to, "...smaller numbers of subjects and to chance" although these are largely one and the same. Perhaps chance was a factor in lowering the ORs. But the sensitivity analyses appear to indicate that the process of imputation (replacing a subject's missing value for an exposure index with an estimate derived from similar subjects) exerted a different effect on the data of the controls than on the data of the cases. This effect was both substantial and systematic.

Finally, the reported findings for the indices that are based in part on the exposure experiment, particularly the estimates of peak exposure, are so unreliable as to be uninterpretable.

VI. Statistical Analyses

Our letter indicated that the most reliable data (referring to the ORs in Table 4 of the ES) on exposure-response were not tested for statistical significance. The authors of the ES responded by quoting one sentence from their paper stating that "Tests of trend for the categorical variables were based on the estimated slope of the original continuous variable (Wald test)". The ES provides no slope parameter estimates from the unconditional logistic regression analysis of the continuous exposure values and no indication that the underlying assumptions and fit of the model were evaluated. If we assume that the underlying assumption of linearity was met (i.e., that the logit of the probability of becoming a case (logit p) was a linear function of the continuous exposure), the slope parameter estimates could have been estimated and used. For example, the increases in the OR per unit increase of exposure could have been provided.

The ES presents p-values that resulted from the continuous (unconditional logistic regression) analyses. These p-values are somewhat misleadingly placed in juxtaposition to the ORs derived from the categorical analyses presented in Tables 3 and 4. But these categorical ORs, which comprise the only descriptive results in the ES, have not themselves been tested for statistical significance in either table. To illustrate: consider the results in Table 4 for "peak" exposure and myeloid leukemia. The ORs for increasing levels of peak exposure to formaldehyde are 2.9, 2.0, 2.9 and the seemingly related p-value (excluding the crucial reference group with an OR=1.0) is an *inverse,* non-significant -0.778. When the reference group is included the "continuous variable" test gives a positive result (p=0.036); but this test is not based on the categorical ORs. In any case, and no matter how it is assessed, there is no exposure-response relationship among the critical groups, the exposed subjects themselves.

The combination just mentioned - the statistically non-significant inverse trend and the statistically significant positive trend seen in the continuous data and an essentially flat pattern of categorical ORs among exposed subjects could have arisen because of an underlying *non-linear* relationship between logit p and continuous exposure. For example, if the underlying logit p vs. continuous exposure relationship was quadratic with a strong linear component, the continuous model (including unexposed subjects) could yield a statistically significant positive slope even with no trend in the ORs among exposed subjects.

A more informative approach to unconditional logistic regression modeling would be an evaluation of the linearity assumption of the continuous model (i.e., linearity of logit p vs. continuous exposure). If linearity is met, the corresponding slope estimates and confidence intervals could then be evaluated. A check of the linearity assumption could be made, for example, using a scatterplot smoothing procedure (e.g., lowess) or by using linear splines of the logit p vs. the continuous exposure.

In addition to the continuous model estimates, the authors should present trend tests based on the midpoints (or scores) of the categorical data that they presented. The arbitrariness of the categories notwithstanding, trend tests based on their midpoints (which

also require underlying linearity, here with respect to logit p vs. midpoint score) would directly reflect the apparent pattern of categorical ORs presented in the tables and avoid possible misinterpretation or confusion about trends.

Full Disclosure – *This document was prepared without financial support. The authors previously have received support from the Formaldehyde Council, Inc. More recently, Drs. Cole, Mandel and Marsh have received support from Hexion Specialty Chemicals, Inc.*

References

1. Hauptmann M, Stewart P, Lubin J, et al: Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. ***J Natl Cancer Inst***. 101(24):1696-1708, 2009.

2. Stewart P, Herrick R, Feigley C, et al: Study design for assessing exposures of embalmers for a case-control study. Part I. Monitoring results. ***Appl Occup Environ Hyg.*** 7(8):532-540, 1992.

3. Cole P, Adami H-O, Trichopoulos D, Mandel J. Lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. ***J Natl Cancer Inst***, online.

4. Hauptmann M, Stewart P, Lubin J, et al. Response. ***J Natl Cancer Inst***, online.

5. Hayes R, Blair A, Stewart P, et al: Mortality of U.S. embalmers and funeral directors. ***Am J Ind Med***.18(6);641-652, 1990

6. Walrath J, Fraumeni J: Mortality patterns among embalmers. ***Int J Cancer***. 31(4):407-411, 1983.

7. Walrath J, Fraumeni J: Cancer and other causes of death among embalmers. ***Cancer Res***. 44(10):4638-4641, 1984.

8. Cole P, Adami H-O, Trichopoulos D, Mandel J. Formaldehyde and lymphohematopoietic cancers: A review of two recent studies. ***Regul Toxicol Pharmacol.*** 58:161-166, 2010.

9. Beane Freeman L, Blair A, Lubin J, et al: Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: update of the NCI cohort. ***J Natl Cancer Inst.*** 101(10):751-761, 2009.

Attachment C:
C1: Cole et al. (2010): Letter to JNCI

CORRESPONDENCE

Re: Mortality From Lymphohematopoietic Malignancies and Brain Cancer Among Embalmers Exposed to Formaldehyde

The Journal recently published an epidemiological study by Hauptmann et al. (1) that focused on formaldehyde exposure and mortality from lymphohematopoietic cancer among embalmers. It concludes that duration of embalming practice is " . . . associated with statistically significantly increased risk of mortality from myeloid leukemia."

The study is based on 6808 embalmers who died from 1960 to 1985 and were included in three earlier proportional mortality ratio studies (2–4). No proportional mortality ratios are presented in the new study, and so we calculated several. Comparison was with deaths among white men who were aged 25 years or older in the United States in 1979, the likely median year of deaths among the embalmers. For all lymphohematopoietic cancers, the proportional mortality ratio was 90 (95% confidence interval [CI] = 76 to 106) and, for the myeloid leukemias, it was 108 (95% CI = 70 to 156), which was based on 29 deaths. The proportional mortality ratio of 128 (95% CI = 35 to 328) for nasopharyngeal cancer was based on four deaths among the embalmers. Proportional mortality ratio studies have limitations and may be misleading. However, we suggest that their principal limitation (self-selection into the study group) is minimal in the embalmers study because of the geographic range (virtually nationwide) and time span of their starting employment (50 or more years).

Interviews with coworkers and next of kin of the decedent embalmers were conducted in 1990–1992. These interviews related to occupational exposures that occurred from the 1920s through 1985. Participation rates are not provided for these interviews. Six indices of formaldehyde exposure were derived for each subject from the interviews and from the findings of an exposure assessment experiment (5). The new study includes no measurement of formaldehyde levels.

The study conducted a case–control analysis of 168 case subjects who died of a lymphohematopoietic cancer and 265 control subjects. Results are presented as odds ratios for three levels of exposure (compared with nonexposed) and the related trend tests for each of the six exposure indices. The analyses were done twice for myeloid leukemias. The results of the first set of analyses presented were implausible, with odds ratios averaging 11, because the referent group included only one case subject who had died of myeloid leukemia. We agree with the authors of the embalmers study that the second set of analyses, with a referent group of five case subjects and odds ratios averaging approximately 2.5, ". . . represent more reliable estimates. . . ."

Surprisingly, the more reliable exposure–response analyses are not accompanied by their attendant *P* values. Even more perplexing is that they were accompanied by the *P* values obtained from the less reliable data. We did not attempt to estimate *P* values from the more reliable analyses. But, viewing these data, we note minimal trends, at most, in the odds ratios for all six indices of exposure among exposed subjects.

We are left with a study that is described as positive for a formaldehyde–myeloid leukemia association among embalmers but which provides little evidence of an overall excess of myeloid leukemia among them and whose most reliable data on exposure–response relationships were not tested for statistical significance.

PHILIP COLE
HANS OLOV ADAMI
DIMITRIOS TRICHOPOULOS
JACK S. MANDEL

References

1. Hauptmann M, Stewart P, Lubin J, et al. Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. *J Natl Cancer Inst.* 2009;101(24):1696–1708.
2. Walrath J, Fraumeni J. Mortality patterns among embalmers. *Int J Cancer.* 1983;31(4):407–411.
3. Walrath J, Fraumeni J. Cancer and other causes of death among embalmers. *Cancer Res.* 1984;44(10):4638–4641.
4. Hayes R, Blair A, Stewart P, Herrick R, Mahar H. Mortality of U.S. embalmers and funeral directors. *Am J Ind Med.* 1990;18(6):641–652.
5. Stewart P, Herrick R, Feigley C, et al. Study design for assessing exposures of embalmers for a case-control study, part I: monitoring results. *Appl Occup Environ Hyg.* 1992;7(8):532–540.

Notes

Affiliations of authors: Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL (PC); Department of Epidemiology, Harvard School of Public Health, Boston, MA (HOA, DT); Department of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada (JM).

Correspondence to: Jack S. Mandel, PhD, MPH, Department of Epidemiology, Dalla Lana School of Public Health, University of Toronto, 155 College St, Rm 638, Toronto, ON, Canada M5T 3M7 (e-mail: jack.mandel@utoronto.ca).

Drs J. S. Mandel, D. Trichpoulos, H. O. Adami, and P. Cole have received grant support from the Formaldehyde Council, Inc, for a different project. Dr. P. Cole has served as a consultant to producers and/or users of formaldehyde.

DOI: 10.1093/jnci/djq332

C2: Hauptmann et al. (2010): Response

CORRESPONDENCE

Response

In three surveys in the funeral industry, the National Cancer Institute reported elevated proportional mortality ratios for leukemia of 140 (1), 175 (2), and 152 (3). However, the relationship of mortality with work practices in the industry was unknown because no occupational histories had been ascertained for these deceased members of the industry in those studies. Our new analytic case–control study (4) addressed this shortcoming by interviewing family members (next of kin) and coworkers of deceased case and control subjects from those earlier surveys (1–3) to determine lifetime work histories and estimate exposure to formaldehyde. On the basis of internal relative risk comparisons that were adjusted for potential confounders, we showed that the excess mortality from leukemia in this industry was specific for myeloid leukemia and was, indeed, associated with embalming practice (for increasing number of years of embalming, P for trend = .020) and exposure to formaldehyde (for increasing peak formaldehyde exposure, P for trend = .036) (4).

The crude proportional mortality ratios calculated by Cole et al. from data reported in our article provide no additional useful information on this topic. The data in our article were insufficient for standard proportional mortality ratio calculations, and Cole's proportional mortality ratios differ, indeed, from the appropriately calculated elevated proportional mortality ratios in our earlier surveys (1–3). As those surveys have shown, proportional mortality ratio studies can sometimes be useful, particularly in the early stages of an investigation. However, among the various measures of relative risk, the proportional mortality ratio has the most limitations and biases. There has been widespread agreement in the epidemiological community for some time that, when available, internal comparisons that are based on individualized exposure assessment provide a much more useful estimate of relative risk than an external comparison, particularly when the external comparison is based on mortality proportions (5,6).

On other issues raised by Cole et al., the interview response rates were included in our article (4): "During 1990–1992, we interviewed at least one next of kin for 220 (96%) of the 228 eligible case subjects and for 265 (94%) of 282 eligible control subjects." The information obtained was quite complete: "Reported work histories covered 18 534.5 (97%) of the 19 104 person-years between the start of the first and the end of the last reported job. Virtually all reported jobs (99.7% of person-years) were characterized by study respondents as being in a funeral home or not, and, for all jobs in funeral homes, it was reported whether the job included embalming." (4). Regarding the question about statistical tests, our trend tests are unaffected by the choice of baseline category for the categorical analysis. As we explained, "Tests of trend for categorical variables were based on the estimated slope of the original continuous variable (Wald test)." (4). That is, we used actual values rather than arbitrary groupings.

M. HAUPTMANN
P. A. STEWART
J. H. LUBIN
L. E. BEANE FREEMAN
R. W. HORNUNG
R. F. HERRICK
R. N. HOOVER
J. F. FRAUMENI Jr
A. BLAIR
R. B. HAYES

References

1. Walrath J, Fraumeni JF Jr. Mortality patterns among embalmers. *Int J Cancer.* 1983;31(4):407–411.
2. Walrath J, Fraumeni JF Jr. Cancer and other causes of death among embalmers. *Cancer Res.* 1984;44(10):4638–4641.
3. Hayes RB, Blair A, Stewart PA, Herrick RF, Mahar H. Mortality of U.S. embalmers and funeral directors. *Am J Ind Med.* 1990;18(6):641–652.
4. Hauptmann M, Stewart PA, Lubin JH, et al. Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde in the funeral industry. *J Natl Cancer Inst.* 2009;101(24):1696–1708.
5. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume II—The Design and Analysis of Cohort Studies. Number 82 in IARC Scientific Publications.* Oxford, UK: Oxford University Press; 1987.
6. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2008.

Notes

Affiliations of authors: Biostatistics Branch (MH, JHL), Occupational and Environmental Epidemiology Branch (PAS, LEBF, RNH, AB), Office of the Director (JFF), Epidemiology and Biostatistics Program (RNH), Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD (MH, PAS, JHL, LEBF, RNH, JFF, AB, RBH); Stewart Exposure Assessments, LLC, Arlington, VA (PAS); Cincinnati Children's Hospital Medical Center, Cincinnati, OH (RWH); Department of Environmental Health, Harvard School of Public Health, Boston, MA (RFH); Department of Environmental Medicine, NYU Medical Center, New York, NY (RBH).

Correspondence to: Laura E. Beane Freeman, PhD, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd, Bethesda, MD 20892 (e-mail: freemala@mail.nih.gov).

Present address: Department of Bioinformatics and Statistics, Netherlands Cancer Institute, Amsterdam, the Netherlands (M. Hauptmann).

DOI: 10.1093/jnci/djq333

Published by Oxford University Press 2010.

Attachment D:
Letter to Dr. Birnbaum (October 29, 2010)
Additonal Data and Analyses of the Zhang et al. (2010) Study

October 29, 2010

Linda S. Birnbaum, Ph.D., D.A.B.T., A.T.S.
Director
National Institute of Environmental Health Sciences
& National Toxicology Program
P.O. Box 12233
Research Triangle Park, North Carolina 27709

Re: NTP Evaluation and Classification of Formaldehyde
for the 12th Report on Carcinogens

Dear Director Birnbaum:

This responds to your October 5, 2010 letter to each of us regarding our September 7, 2010 letter to Dr. Francis Collins, Director of the National Institutes of Health, a copy of which is enclosed (as Attachment A). Your October 5 letter enclosed "an informational document" on the evaluation of formaldehyde for the 12th Report on Carcinogens (RoC) that states:

> *The NTP is currently completing the review process for the candidate* substances under consideration for listing in the 12th RoC. *Once completed, the NTP will consider* the input from the BSC peer review, public comments, and *any new, relevant scientific information on formaldehyde in finalizing its recommendation on the listing status of formaldehyde, and will prepare the draft RoC*. The NTP Director will share the draft RoC with the NTP Executive Committee and then transmit it to the Secretary, Health and Human Services, for review and approval. Once approved, the Secretary will transmit the 12th RoC to Congress and it will be released to the public.

(Italics added.)

Thus, according to your letter and informational document, the NTP will consider "new, relevant scientific information on formaldehyde" before completing its review process for

formaldehyde as a candidate for listing in the 12th RoC. This letter provides new, critical information that is directly related to the NTPs deliberations on formaldehyde and was received after the NTP's public comment period for the 12th ROC ended in June. Specifically, we have obtained (through a FOIA request) data that were used in the report by Zhang et al. (2010). This includes very important information which bears directly on the validity of this study's conclusions and its subsequent use in NTP's assessment of formaldehyde as a human leukemogen.

Analyses based on individual worker data from the Zhang et al. (2010) study.

The NTP has relied heavily upon the Zhang et al. (2010) study in reaching its draft conclusion that formaldehyde is a known human leukemogen. We conducted a detailed analysis of this study, specifically including the underlying data received in July from the National Cancer Institute (NCI), the study sponsor, in response to an earlier request pursuant to the FOIA[1]. NTP did not consider these worker-specific data in NTP's evaluation of the Zhang et al. (2010) study. Our analyses have revealed numerous methodological limitations, unwarranted underlying biological assumptions, and evidence that the chromosome aneuploidy, if indeed present and which forms the basis of the authors' conclusions, most likely arose *in vitro* and is unrelated to formaldehyde exposure in the work place.

Specifically, the aneuploidy data for individual workers do not support injury to "stem" cells and cannot be due to events occurring *in vivo* as a result of exposure of workers to formaldehyde (Table 1). Serious problems include the following:

- The raw data clearly show that the methodology described in Zhang et al. (2010) was not followed (Table 1) -- that is, "a minimum of 150 cells per subject" were not scored.
    - Only 1 of 10 exposed worker and only 4 in 14 non-exposed workers had greater than or equal to 150 cells scored.
    - For the remaining 17 cases, the total number of cells scored ranged from 18 to 140 (see bullet on additional statistical analysis below).
- The method used by Zhang et al. (2010) requires a minimum of six doublings over a 14-day period in order to measure a clonal event occurring in a CFU-GM cell inoculated into

[1] We have submitted a letter to the National Academies of Science providing specific criticisms of the conduct and interpretation of the Zhang et al. (2010) data based on the information received in reference to the FOIA request (Attachment B).

cell culture (Stem Cell Technologies). A review of the aneuploidy data for the individual workers showed that the largest number cells scored with either monosomy 7 or trisomy 8 was 20, and that the vast majority of cells with changes in either chromosomes 7 or 8 was less than 10 in either the exposed or non-exposed groups.

- The number of monosomy 7 events detected in exposed workers ranged from 4% to 16% of cells scored, and the percentage of trisomy 8 events ranged from 0% to 2% of cells scored.

- These percentages of aneuploid metaphases (either type) are far too low to have been due to the contributions of aneuploid colonies derived from aberrant cells pre-existing the in vitro assay, i.e. from cells already mutated in vivo. Rather, these percentages should have been greater than 50% of the total metaphases scored for most of the individuals studied (Table 1) if most in vitro progeny from even a single aneuploid colony derived from an in vivo aneuploid progenitor had been recovered, i.e. six cell doublings equals 128 in vitro progeny, half of which should have been in metaphase at termination of the in vitro assay. If in vitro progeny had been recovered from several aneuploid colonies, each derived from an aneuploid cell that had originated in vivo, this percentage of aneuploid metaphases would rise towards 100%.

- Based on the small number of aneuploid metaphases detected and the kinetics of CFU-GM colony formation (Lewis et al. 1994), all of the aneuploid metaphases scored by Zhang et al. most likely occurred in or after the 6^{th} division in culture, arising *in vitro* in the assay system rather than *in vivo* as a result of formaldehyde exposure to workers.

- Additional statistical analyses of the FOIA data have been conducted using the individual worker data (restricted to subjects for whom greater than 80 cells, the average number scored, were examined).

  - Using the same statistical method described by Zhang et al. (2010), no statistically significant difference existed in exposed workers, when compared to unexposed workers, for either monosomy 7 or trisomy 8.

  - Use of Chinese medicine alone for each individual worker (without consideration of formaldehyde exposure) was significantly associated with aneuploidy.

Another serious limitation of the method used by Zhang et al. (2010) became clearer after a review of these individual worker data. Specifically, as noted in Attachment B, Zhang et al. (2010) used an assay that measured, CFU-GM from a peripheral blood cell, and assumed that this cell is a progenitor stem cell causally related to the production of myeloid leukemia. However, CFU-GM is not the appropriate assay for measuring clonal events in the development of myeloid leukemia, specifically, acute myeloid leukemia (Lapidot et al. 1994; Terpstra et al. 1996; Ailles et al. 1999).

- Cells capable of becoming AML-initiating cells (AML-IC) cannot be measured by CFU-GM:
    - CFU-GM cells are endstage cells with limited or no proliferation potential, and are incapable of either self-renewal or the propagation of leukemia in the model system proposed by Zhang et al.
    - Other assays, such as the cobblestone-area forming cells (CAFC) or human severe combined immunodeficient (NOD/SCID) repopulating cells (huSRC), are necessary and appropriate for monitoring clonal development of lesions in AML, while CFU-GMMM and CFU-GM are not.

- Normal human peripheral blood repopulating (stem) cells are vanishingly rare and are not in cell cycle.
    - The underlying hypothesis by Zhang et al. (2010) is that aneusomy occurs in circulating hematopoietic stem cells which, by definition, must be dividing, and must return to the bone marrow.
    - The frequency of circulating huSRC is so rare as not to be measurable, and that attempts to use *un-mobilized* circulating cells in order to reconstitute hematopoiesis in mice and humans has failed.
    - Because of the very limited amount of systemically available exogenous formaldehyde compared to that formed exogenously, such changes, if they occurred at all, more likely would be associated with normal endogenous formaldehyde production.

These observations and analyses conducted using data provided under the FOIA and from other sources, support our conclusion that the Zhang et al. (2010) study presents no convincing evidence for a mechanistic link between exposure to formaldehyde and myeloid leukemia. In

light of this new and relevant information, we request the NTP to thoroughly evaluate and specifically address the important scientific matters contained in this letter and in our September 7 letter to Dr. Collins, prior to NTP's completion of the 12th ROC. Moreover, in light of the current review of the EPA IRIS draft document on the carcinogenicity of formaldehyde by an expert committee of the National Academies of Science, it would be prudent to postpone the inclusion of formaldehyde as a leukemogen in the 12th ROC until it is clear that the NAS expert committee has completed its work, otherwise, a direct and avoidable conflict between NAS and NTP will result.

Thank you for your serious consideration of this new, relevant information and analysis. Please contact Dr. Kenneth Mundt (at 413-256-3556) if you have questions or otherwise wish to discuss these matters. Thank you.

Richard Albertini, MD, PhD
Professor Emeritus, Medicine
Research Professor, Pathology
University of Vermont College of Medicine
Burlington, VT

Philip Cole, MD
Professor Emeritus of Epidemiology
University of Alabama at Birmingham
Birmingham, AL

Richard Irons, PhD
Professor Emeritus
University of Colorado at Denver
Cinpathogen International, LLC
Boulder, CO

Jack S. Mandel PhD, MPH
Chief Science Officer
Exponent
Menlo Park, CA

Gary M. Marsh, Ph.D., F.A.C.E.
Professor of Biostatistics, Epidemiology and Clinical & Translational Science
Director, Center for Occupational Biostatistics & Epidemiology
Department of Biostatistics
Graduate School of Public Health
University of Pittsburg
Pittsburgh, PA

Kenneth A. Mundt, PhD
Principal
ENVIRON
Amherst, MA

Enclosures

REFERENCES

Ailles, L., et al. (1999) Growth characteristics of acute myelogenous leukemia progenitors that initiate malignant hematopoiesis in nonobese diabetic/severe combined immunodeficient mice. Blood 94(5): p. 1761-1772.

Lapidot, T., et al. (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 367: 645-648.

Lewis, J., N. Blackett, and M. Gordon (1994) The kinetics of colony formation by CFU-GM in vitro. British J of Haematology 88:440-442.

Terpstra, W., et al. (1996) Fluorouracil selectively spares acute myeloid leukemia cells with long-term growth abilities in immunodeficient mice and in culture. Blood 88:1944-1950.

Zhang, L; Tang, X; Rothman, N; et al. (2010) Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. Cancer Epidemiology, Biomarkers & Prevention 19(1):80-88.

Table 1
Levels of Monosomy of Chromosome 7 and Trisomy of Chromosome 8 in cells Scored by (Zhang et al. 2010)

Monosomy 7				Trisomy 8			
Exposed		Unexposed		Exposed		Unexposed	
Number Detected	Number Analyzed	Number Detected	Number Analyzed	Number Detected	Number Analyzed	Number Detected	Number Analyzed
11	274	19	288	2	192	2	226
15	132	10	272	4	180	2	215
20	123	10	260	4	173	2	197
4	109	8	163	1	149	1	94
4	101	6	140	0	139	0	91
3	95	2	78	0	108	0	83
9	76	1	71	2	78	0	69
13	61	9	70	2	61	0	67
10	50	4	49	0	53	0	37
6	39	0	24	0	33	0	25
		2	20			0	22
		1	18			0	21

Attachment E: Analyses of the Zhang et al. (2010) study
Letter to the NAS October 4, 2010

Dr. Ellen Mantus
Senior Program Officer for Risk Analysis
National Academy of Sciences
500 Fifth St., N.W.
Washington, D.C. 20001
Email: emantus@nas.edu

RE: **Summary of Critical Concerns Regarding Reliance on Zhang et al, *Cancer, Epidemiology, Biomarkers & Prevention*, Vol. 19(1), pp 80-88 (2010) as Reported in the Draft IRIS Formaldehyde Assessment**

Dear Members of the NAS Committee:

Prior to 2009, agencies such as IARC (2006) had concluded that the results from epidemiological studies, such as Hauptmann et al. (2003), did not provide sufficient evidence to classify formaldehyde as a known human leukemogen. However, recent assessments by IARC, and by staff of NTP and USEPA, have found that there is now sufficient evidence for a causal association with myeloid leukemia (Baan et al. 2009; NTP 2010a, b; USEPA 2010). While epidemiological studies conducted by Beane Freeman et al. (2009) and Hauptmann et al. (2009) were considered[2], in reaching these findings, much weight, and apparently the deciding factor, has been given to the results provided by Zhang et al. (2010).

As described in the remainder of this letter, there are many serious methodological deficiencies, reporting errors and unsupported biological assumptions in the Zhang et al. (2010) study[3], and these render the conclusions reached by Zhang et al. (2010) at best an over-interpretation of their results, and, in most cases a misinterpretation of their results.

Zhang et al. (2010) purport to measure what they refer to as "leukemia-specific" chromosome changes in CFU-GM cells because CFU cells are the target cells for leukemogenesis and are converted to leukemia stem cells in acute myeloid leukemia (AML)." This is an erroneous assumption based on outdated 30 year old theories of leukemogenesis that are not supported by current scientific and medical knowledge:

---

[2] Drs. Cole, Mandel, Marsh and Mundt have submitted to this NAS Committee a critique of the primary epidemiological studies. Their review concluded that there is not a causal association with myeloid leukemia because: 1) no statistically significant association between formaldehyde exposure and myeloid was seen for any dose-metric in the Beane Freeman et al. (2009) study, and, 2) methodological limitations in the Hauptmann et al. study call into question the conclusions drawn by the authors.

[3] Individual data for a specific test reported by Zhang et al. (2010) were obtained from NCI in response to a Freedom of Information Act (FOIA) request, and those data provide compelling evidence of these limitations.

A. CFU-GM is not the appropriate assay for measuring clonal events in the development of AML.

- **Cells capable of becoming AML-initiating cells (AML-IC) cannot be measured by CFU-GM.**

The maturation hierarchy of AMLs is analogous to that of cells in normal hematopoiesis. Even in AML, the vast majority of leukemia blasts are endstage cells with limited or no proliferation potential, and they are incapable of self-renewal or propagation of the leukemia. An assay that identifies AML-IC, and that would be useful for demonstrating a clonal lesion in the development of an AML, necessarily must measure a cell capable of self-renewal. In the modern era, the advent of xenograft assays and recombinant cytokine technology has led to a general understanding that AML-IC are less mature than colony-forming cells (e.g. CFU-GM). AML-IC, specifically including those arising in AMLs with aneusomy (e.g. +8, -7), correspond to long-term culture initiating cells (LTC-IC), cobblestone-area forming cells (CAFC), or human severe combined immunodeficient (NOD/SCID) repopulating cells (huSRC)[1-3]. The studies reported by Lapidot et al. (1994), Terpstra et al. (1996) and Ailles et al. (1999) and other studies have demonstrated that AML-IC can be measured using in vitro assays, and that their self-renewal capability is confirmed in vivo. However, AML-IC cannot be measured using CFU-GM or other CFU assays, because such assays do not measure cells with self-renewal capability.

Conditions for culture of huCFU-GM, including those described by Zhang et al (2010), call for incubation for 14 days in methylcellulose with GM-CSF. In contrast, all of the assays for AML-IC require either long term culture (~ 6 weeks) under complex conditions, or transplantation into NOD/SCID mice, in order to identify and measure their proliferative potential. These early cells do not produce differentiated progeny (i.e. colonies) in CFU assays in less than 3 weeks, and they cannot survive under the conditions used for CFU assays, including methylcellulose or agar medium + GM-CSF for CFU-GM[4]. For example, selection and enrichment of AML-IC, using in vitro incubation with agents such as 5-fluorouracil (5-FU), although they do not result in any CFU-GM [2], they do result in enrichment of cells that form CAFC (6 week cultures), and are capable of leukemic transplantation in NOD/SCID mice. Consequently, CAFC and NOD/SCID assays are appropriate for monitoring clonal development of lesions in AML, while CFU-GEMM and CFU-GM are not.

- **Normal human peripheral blood repopulating (stem) cells are vanishingly rare and are not in cell cycle.**

Central to the mechanism of action proposed by Zhang et al. (2010) is the premise that aneusomy occurs in circulating hematopoietic stem cells, which by definition must be dividing, and then return to the bone marrow, the tissue of origin for myeloid leukemias. Although it is

impossible to prove a negative, this premise is unsupported. In normal untreated individuals who have not undergone cytokine therapy to mobilize primitive repopulating cells (i.e. stem cells), the frequency of circulating huSRC is so rare as to not be readily measurable [5]. Data on the frequency of circulating stem cells that return to the bone marrow is virtually non-existent, largely because they are so rare that early attempts to use un-mobilized circulating cells to reconstitute hematopoiesis in humans failed ([6-7]). Similarly, the vast majority of circulating repopulating cells in humans are in G_0, and G_0-G_1 progression actually results in depletion of repopulating capacity [8].

B. The methodology used by Zhang et al. (2010) to measure FISH in CFU-GM is seriously inadequate.

- **The reported data do not permit analysis of clonal injury.**

The only advantage of using CFU-GM over liquid culture to measure terminally differentiating myeloid cells is the potential to measure clonal events, i.e. the number of cells containing lesions per colony. However, because Zhang et al. did not report either colony number or the number of colonies scored for fluorescence in situ hybridization (FISH), it is not possible to determine clonal injury.

- **Results of FISH reported by Zhang et al cannot be due to an event occurring in vivo.**

Zhang et al. (2010) report using the CFU-GM protocol published by Stem Cell Technologies, which scores colonies at 14 days containing $\geq$ 40 cells. This requires a minimum of 6 doublings occurring over a 14-day period to measure a clonal event occurring in a CFU-GM progenitor cell inoculated into culture. A review of the Zhang et al. (2010) primary FISH signal data (n's), obtained from NCI through a Freedom of Information Act (FOIA) request, show that the largest n reported for a FISH signal was 20. The vast majority of n's are $\leq$ 10 for both the control and "exposed" groups. Based on the kinetics of CFU-GM colony formation [9], all of these signals most likely occurred in or after the 6th division in culture, and almost certainly do not represent events that could have taken place in vivo.

- **The raw data do not comport with the methodology described in Zhang et al. (2010).**

The authors reported that they counted a minimum of 150 cells for each case (exposed = 10, controls = 12). However , the raw data show that far fewer cells were analyzed in the majority of cases. For monosomy 7, a review of the raw data reveals only 1 exposed and 4 control cases for which 150 cells were, in fact, scored. And for the remaining 17 cases, the total number of cells counted ranged from 18-140. (See Appendix 1 to this letter for the table of cells scored for monosomy 7 and trisomy 8.) FISH assays, including those utilizing the specific probes employed by Zhang et al. (2010), are subject to correction for background/sensitivity errors. Because of statistical limitations inherent in the scoring of FISH assays, a minimum of

200 cells (and for certain probes more) are required to report a result in a clinical setting. Statistically significant differences reported for +8 were 1.21% and 0.32% for exposed (n=10) vs unexposed (n=12) subjects. However, if analysis is limited to cases where even ≥ 100 cells are counted, the percentage with +8 is nearly identical (i.e. 1.04% vs 0.94%, respectively).

- **Cutoff values for FISH in normal individuals are not presented.**

Zhang et al (2010) fail to provide cut-off values for +8 and -7 FISH probes in normal individuals. However, in a previous study they indicated an apparent cutoff value of 0.8% ± 0.1% for +8 in controls. [10]. These values call into question whether their FISH analyses can meaningfully resolve the small differences reported in this study.

- **Aneuploidy increases with time in human cells cultured in vitro.**

It is well established that the frequency of aneuploid cells spontaneously increases with time in culture, with significant differences reported in human lymphocytes as early as 72h [11-13]. Further, increases in the frequency of aneuploidy are not random, and numerical changes involving chromosome 8 occur at a higher rate in culture than for many other chromosomes [14].

- **Validation of the CFU-GM/FISH assay to measure aneuploidy in vivo is inadequate**

Previously, Zhang et al. reported that the frequency of aneuploidy in cultured human lymphocytes scored by FISH is several-fold higher in low- and medium-quality metaphase preparations, versus high-quality metaphase preparations [10]. (See Appendix 2 in which Zhang et al. discuss the reliability of their FISH technique specifically for chromosomes 7 and 8.) Given the extended incubation time for CFU-GM used in this study (14d), and the fact that the authors report they scored all available cells, there is a high probability that low- and medium-quality metaphase preparations were counted, leading to an overestimate of aneuploidy.

Notwithstanding that CFU-GM do not measure a repopulating stem cell population, individual variability, inter-testing variability and standard reference ranges should be characterized in a large normal population prior to the application of this methodology, even for use as a non-specific biomarker of effect.

C. Changes in blood parameters reported by Zhang et al (2010) are not clinically significant.

- **Total counts for WBC, neutrophil, lymphocyte and platelets are well within normal limits for both exposed and unexposed groups.**

- **Statistically significant differences reported between exposed and unexposed groups cannot be attributed to formaldehyde.**

In light of published genetic and regional differences in blood cell counts in Asian populations (see below), insufficient information is provided in Zhang et al. (2010) on the clinical background and origin of exposed individuals (n=43) versus unexposed (n=51) controls to evaluate potential confounding for the small differences between groups reported.

Examples of other potential confounders include:

1) Thalassemia trait.

Although differences are reported for RBC and MCV, Hgb values are identical between exposed and control groups. Analysis of primary data reveals 1 exposed and 4 controls meet criteria (MCV <70fL), suggestive of thalassemia trait. Re-analysis of the data excluding these subjects narrows the difference between MCV and RBC between the two groups, implicating thalassemia trait as a likely confounder.

2) Genetic and regional confounders.

Significant variations in platelet counts in healthy Chinese subjects are known to be influenced by such factors as geographical location, season, and lipid variations [15]. Similarly, genetic polymorphisms have been identified that are associated with significant differences in neutrophil counts in Asian populations [16]. Other influences such as nutrition and Chinese medicine have not been appropriately addressed, and these influences could easily explain the minor variations in CBC parameters that Zhang et al. (2010) instead have attributed to formaldehyde.

- **"Differences" in CFU-GM colonies are not statistically significant.**

Zhang et al describe a "20% decrease in CFU-GM colonies" between exposed [Mean = 7.26%; Range: 1.32-21.38%] and unexposed [Mean = 9.03%; Range:0.84-22.88] groups, which they suggest is due to a toxic effect of formaldehyde even though the results are not statistically significant (p= 0.10). <u>The reported differences in the averages are likely meaningless in view of the fact that there is over a 20-fold variation in the number of colonies in both exposed and unexposed groups.</u>

<u>D. There are no validated mechanisms for the pathogenesis of AML with aneusomy.</u>

- **Neither "Myeloid leukemia" nor AML is a single disease.**

The WHO classification of hematopoietic malignancies stratifies myeloid neoplasms according to broad categories (<u>e.g</u>. chronic myeloproliferative neoplasms, including chronic myelogenous leukemia (CML); myelodysplastic/myeloproliferative diseases (MDS/MPD); myelodysplastic syndromes (MDS); and acute myeloid leukemia (AML)). Insofar as is possible, distinct diseases are defined within each category based on morphology, immunophenotype, genetic abnormalities, clinical features and etiology [17]. Within the category of AML there are several

subdivisions comprised of approximately a total of 20 different disease entities [17-18]. Therefore *a priori* it is not appropriate to consider all myeloid leukemias, or even major subgroups of AML, as a single disease [19].

- **The relevance of monosomy 7 or trisomy 8 as markers for formaldehyde exposure is not established.**

AML, with specific genetic mutations and recurrent cytogenetic abnormalities, are classified on the basis of structural chromosome rearrangements, thereby resulting in chimeric proteins or molecular genetic changes that correlate with alterations in gene expression (gene activation or inactivation). The pathogenesis of these entities, together with their prognostic significance, are well defined and form the basis for classification of CML and AML with reoccurring cytogenetic abnormalities [17].

Zhang et al. (2010) refer to their identification of monosomy 7 and trisomy 8 as "Leukemia-Specific" chromosome changes. However, there is no mention in Zhang et al. (2010) regarding analysis for potential aneuploidy of other chromosomes. Because there are no chromosome studies available for the leukemia cases reported in either controls or in exposed individuals in the available epidemiologic studies, it is not possible to ascertain whether monosomy 7 or trisomy 8 are relevant markers to study in formaldehyde exposed workers.

- **Benzene-induced MDS/AML is not a positive control for aneusomy occurring in the pathogenesis of hematopoietic neoplasms.**

Over the past 20 years, hypotheses for a role for aneuploidy in the development of benzene-induced hematopoietic neoplasms have been the subject of numerous studies, including several authored by one of us. However, these hypotheses are not supported by the results of recent epidemiology studies employing state-of-the-art molecular and cytogenetic methods. Although current studies confirm previous observations that benzene plays a causal role in development of subtypes of MDS and AML, those studies do not provide any evidence for aneusomy in the development of these diseases [20-21] .

In conclusion, based on the fundamental biological misconceptions, methodological deficiencies and inaccuracies described above, we believe that the Zhang et al. (2010) study cannot be considered reliable, and it should not be used as a basis to confirm or suggest a relationship between formaldehyde and AML or any other leukemia.

Sincerely,

Richard D. Irons, MT, PhD, DABT
Professor Emeritus of Pathology and Toxicology
University of Colorado at Denver
Cinpathogen International, LLC
Boulder, CO

Richard J. Albertini, MD, PhD
Professor Emeritus, Medicine
Research Professor, Pathology
University of Vermont College of Medicine
Burlington, VT

Michael J. Thirman, MD
Associate Professor of Medicine
Director of Leukemia Biology Hematology/Oncology
University of Chicago

Annette M. Shipp, PhD
Principal
ENVIRON International Corporation
Monroe, LA

References

1. Lapidot, T., et al., *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice.* Nature 1994. **367**: p. 645-648.
2. Terpstra, W., et al., *Fluorouracil selectively spares acute myeloid leukemia cells with long-term growth abilities in immunodeficient mice and in culture.* Blood, 1996. **88**: p. 1944-1950.
3. Ailles, L., et al., *Growth characteristics of acute myelogenous leukemia progenitors that initiate malignant hematopoiesis in nonobese diabetic/severe combined immunodeficient mice.* Blood, 1999. **94**(5): p. 1761-1772.
4. Coulombel, L., *Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays.* Oncogene, 2004. **23**: p. 7210-7222.
5. Wang, J., M. Doedens, and J. Dick, *Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay.* Blood, 1997. **89**(11): p. 3919-3924.
6. Abrams, R., et al., *Result of attempted hematopoietic reconstitution using isologous, peripheral blood mononuclear cells: a case report.* Blood, 1980. **56**(3): p. 516-520.
7. Hershko, C., et al., *Cure of aplastic anemia in paroxysmal nocturnal haemoglobinuria by marrow transfusion from identical twin: failure of peripheral blood-leukocyte transfusion to correct marrow aplasia.* Lancet, 1979. **1**: p. 945-947.
8. Gothot, A., et al., *Cell cycle-related changes in repopulating capacity of human mobilized peripheral blood CD34+ cells in non-obese diabetic/severe combined immune-deficient mice.* Blood, 1998. **92**(8): p. 2641-2649.
9. Lewis, J., N. Blackett, and M. Gordon, *The kinetics of colony formation by CFU-GM in vitro.* British J of Haematology, 1994. **88**: p. 440-442.
10. Zhang, L., et al., *Benzene increases aneuploidy in the lymphocytes of exposed workers: a comparison of data obtained by fluorescence in situ hybridization in interphase and metaphase cells.* Environmental & Molecular Mutagenesis, 1999. **34**(4): p. 260-8.
11. Richard, F., et al., *Aneuploidy in human lymphocytes: an extensive study of eight individuals of various ages.* Mutation Research, 1993. **295**: p. 71-80.
12. Sherwood, S., et al., *Defining cellular senescence in IMR-90 cells: A flow cytometric analysis.* Proc Natl Acad Sci, USA, 1988. **85**: p. 9086-9090.
13. Zotova, N., et al., *Spontaneous aneuploidy level in blood cells of fertile females.* Cell and Tissue Biology, 2009. **51**(7): p. 585-591.
14. Mukherjee, A.B. and S. Thomas, *A longitudinal study of human age-related chromosomeal analysis in skin fibroblasts.* Exptl Cell Research, 1997. **235**: p. 161-169.

15. Peng, L., et al., *Effects of biological variations on platelet count in healthy subjects in China.* Thromb. Haemost., 2004. **91**: p. 367-372.
16. Okada, Y., et al., *Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count.* Human Molecular Genetics, 2010. **19**(10): p. 2079-2085.
17. Vardiman, J., et al., *Introduction and overview of the classification of myeloid neoplasms.*, in *WHO Classification of Tumours of the Haematoipoietic and Lymphoid Tissues*, S. Swerdlow, et al., Editors. 2008, International Agency for Research on Cancer (IARC): Lyon. p. 18-30.
18. Harris, N., et al., *WHO classification of tumours of haematopoietic and lymphoid tissues: Introduction*, in *World Health Organization Classification of Tumours*, E. Jaffee, et al, Editor. 2001, IARC Press: Lyon, France. p. 12-13.
19. Appelbaum, F., et al., *Acute myeloid leukemia.* Hematology, 2001. **2001**: p. 62-86.
20. Irons, R.D., et al., *Integrating WHO 2001-2008 criteria for the diagnosis of Myelodysplastic Syndrome (MDS): a case-case analysis of benzene exposure.* Chem Biol Interact, 2010. **184**: p. 30-38.
21. Wong, O., et al., *A hospital-based case-control study of acute myeloid leukemia in Shanghai: analysis of environmental and occupational risk factors by subtypes of the WHO classification.* Chemico-Biological Interactions, 2010. **184**: p. 112-128.

Appendix 1
Levels of Monosomy of Chromosome 7 and Trisomy of Chromosome 8 in cells Scored by (Zhang et al. 2010)

Monosomy 7				Trisomy 8			
Exposed		Unexposed		Exposed		Unexposed	
Number Detected	Number Analyzed	Number Detected	Number Analyzed	Number Detected	Number Analyzed	Number Detected	Number Analyzed
11	274	19	288	2	192	2	226
15	132	10	272	4	180	2	215
20	123	10	260	4	173	2	197
4	109	8	163	1	149	1	94
4	101	6	140	0	139	0	91
3	95	2	78	0	108	0	83
9	76	1	71	2	78	0	69
13	61	9	70	2	61	0	67
10	50	4	49	0	53	0	37
6	39	0	24	0	33	0	25
		2	20			0	22
		1	18			0	21

Appendix 2
Reproduced from Zhang et al. (1999), p. 266 [10]

"One issue that must be addressed is the apparently high rates of monosomy (cells with one hybridization signal) and trisomy (cells with three hybridization signals) reported in metaphase cells in the present study compared with rates obtained by classical cytogenetics. The rates of apparent monosomy and trisomy most likely result from the fact that we examined all scorable metaphases on the slides, as previously defined [Zhang et al, 1998b], rather than just the 50 best. This approach accounts for the differences between our data and the much lower numbers generated by conventional analysis of a limited number of high quality metaphase spreads. When we reexamined around 50 of the best metaphase spreads by FISH in three subjects with high aneuploidy rates, we detected very few aneuploid cells. In fact, on average, the rate of aneuploidy was several-fold lower in the best 50 metaphase spreads compared with those of lower quality (data not shown). Therefore, using FISH to analyze only the best quality spreads would have yielded a lower aneusomy rate comparable to rates obtained by classical cytogenetics. The high values we report here and elsewhere result from our scoring poor- and medium- as well as high quality spreads. This approach appears to increase our power to detect chromosomal damage in exposed populations, but does not allow for ready comparison of our data with measurements of aneuploidy by conventional cytogenetics or even by FISH in the best metaphase spreads."